



## 記事まとめサイト できました!

過去の記事もまとめて確認頂けます<br/>
ぜひ解析のご参考に





Geneious Prime でシークエンス解析

ンス解析 **gene**ious prime

第10回 アンプリコンメタゲノミクス(その2)

メタゲノミクスは、環境サンプルから直接回収された遺伝物質の研究です。前回(<u>第 9 回 アンプリコンメタゲノミクス(その1)</u>)から引き続き、発酵プロセスに関連する細菌群をプロファイルするため、自然発酵したザワークラウトから PCR 増幅された 16S rRNA 遺伝子配列を解析する手法をご紹介します。

今回は de novo アセンブラを使用してリードを OTU (Operational Taxonomic Unit)にクラスタリングする手法についてです。

NGS のアンプリコンデータセットは通常数百万リードからなるため、これらのすべての配列を分類のために BLAST することは現実的ではありません。そこで、リードを類似性でクラスタリングして OTU を定義し、各 OTU の代表的な配列を BLAST します。この BLAST の結果を Sequence Classifier プラグインで分類対象データベースとして利用することで、すべてのリードセットの生物多様性を定量化することができます。

de novo アセンブリを行い、すべての近縁な配列をそれぞれのコンティグにクラスタリングします。各コンティグから得られるコンセンサス配列が OTU を表すことになります。この作業でデータセットが大幅に簡素化され、その後の batch-BLAST に使用するデータセットを減らすことができます。

前回作成した、トリム、マージ、length フィルタリング済みのリードセット(SRR7140083\_50000 (trimmed) (merged) - length 150-260) を選択し、Align/Assemble → De novo assembly に進みます。一度 Sensitivity: を Low sensitivity/Fastest に設定して厳密な設定をロードした後、再度 Sensitivity: を Custom Sensitivity に変更して、以下のように設定を力スタマイズします。これで各コンティグが近縁な配列のみから構成されるようになります。

De Novo Assemble	
Data	
Assemble by: 1st 0 part of name, separated by - (Hyphen)	
Assemble each sequence list separately  Use 100 🔾 % of data. Suitable for genome size between 75 KB and 303 KB.	
Assembler: Geneious ② ?  Sensitivity: Custom Sensitivity ② ?	
	be set up or changed using Sequence > Set Paired Reads
Trim Before Assembly—	Results
	Assembly Name {Reads Name} Assembly
Use existing trim regions	Save assembly report
Remove existing trim regions from sequences	Save list of unused reads
Options	☑ Save in sub-folder
O Do not trim	Save contigs ( Maximum 1,000 🗘 )
	☑ Save consensus sequences Options
Advanced -	
Don't merge variants with coverage over appro	eximately 6 0 Merge homopolymer variants
Circularize contigs of 3 0 or more sequen	nces, if ends match Produce scaffolds
☑ Allow Gaps Maximum Per Read: 1 0 %	
☑ Minimum Overlap: 100 🗘	✓ Minimum Overlap Identity: 98 🕻 %
Word Length: 24 0	Index Word Length: 14 🗘
☑ Ignore words repeated more than 200 🕻 ti	imes Reanalyze threshold: 16 C
Maximum Mismatches Per Read: 2 0 %	6 Maximum Ambiguity: 4 0
Low Memory Use 2 3 4 5 Speed	Only use paired hits during assembly
☆ Fewer Options	Cancel
A react options	Called

上記はデモデータでの設定になりますので、ご自身のケースに合わせて最適化していただく必要があります。 もしデータセットのリードのクオリティが低かったり、属レベルでの解像度の解析で十分だったりする場合などは、Minimum Overlap Identity の設定を下げ、Maximum Mismatches Per Read の設定を上げることで、クラスタリングが厳密でなくなり、OTU がある程度ばらつきのある配列で構成されるようになるため、OTU の数を減らすことができます。

OK をクリックすると、de novo アセンブラが起動します。 de novo アセンブリ処理には 5~10 分程度かかる場合があります(デモデータの場合)。完了するとフォルダの中にアセンブリレポートが入った新しいフォルダが表示されます。このフォルダを選択しアセンブルレポートを表示します。

上記の設定で、デモデータでは60ほどのコンティグと未使用リードが作成されるかと思います。各コンティグのコンセンサス配列は Consensus Sequences リストに、未使用のリードは Unused Reads リストに表示されています。

次回は OTU を batch-BLAST する手法をご紹介する予定です。

Geneious 製品概要についてはこちら